استكشاف وتقدير القيم الشاذة في بعض النماذج اللاخطية*

أ.م.د.حميد ناصر الفتال* م.د.انكين انترانيك هايك**

المستخلص:

ان وجود القيم الشاذة ضمن مجموعة من البيانات يؤثر بشكل كبير على نتائج التحليل الاحصائي للبيانات، من هنا تظهر اهمية تشخيص القيم الشاذة كخطوة اولى ومهمة في عملية التحليل واتخاذ القرار. لذا فان تشخيص الشواذ يمثل واحداً من الاهداف العامة في عملية تحليل البيانات، وتعتبر تقدير القيم الشاذة بعد تشخيصها للأنهاء شذوذها واعادتها الى البيانات كقيم بديلة كخطوة ثانية ومهمة في عملية تحليل البيانات لتطبيق التحليل الأحصائي المناسب .

يمكن تلخيص هدف البحث في نقطتين، الاول يضم دراسة بعض نماذج الانحدار اللاخطية وتأثير وجودها على مقدرات تلك النماذج، الثاني اقتراح طرق جديدة لتقدير المشاهدات الشاذة في بعض النماذج اللاخطية مع اجراء مقارنات بين طرق التقدير المقترحة والطرق الموجودة فعلاً باستخدام متوسط مربعات الخطأ MSE ومعامل التحديد R^2 .

Abstract

* استاذ مساعد/ الجامعة المستنصرية/كلية الادارة والاقتصاد/قسم الاحصاء

* * استاذ مساعد/ الجامعة المستنصرية/كلية الادارة والاقتصاد/قسم الاحصاء

مقبول للنشر بتاريخ 2009/9/9

* مستل من اطروحة د. آنكين أنترانيك

The Appearance of the outlier Values in the set of data may effect on the result of the statistical Analysis of data, from this point of view appears the Importance of detecting the outlier values in any set of data before study. The object of this thesis can be explain in two points. Firstly it is to study some non-linear regression models and its influences appearance on the estimators of that models. Secondly suggest some new methods of estimating the outliers in some non-linear models and comparing between these methods by using the measures MSE and R^2.

1. المقدمة :

يعد تحليل الانحدار من التحليلات الاحصائية المهمة التي استخدمها بصورة واسعة الباحثون في مواقع الانتاج أو الخدمات وان اجراء اى تحليل في النواحي العملية يعتمد بالدرجة الاساس على أختيار مجموعة من البيانات .أن سلامة هذه البيانات وتنقيتها من القيم الشاذة مهمة ضرورية لسلامة النتائج والاستنتاجات المبنية عليها،فالبدايات الاولى لموضوع القيم الشاذة كانت تتعامل مع مثل هذا النوع من المشاهدات بان تحذف ويجرى التحليل على بقية المشاهدات , وحيث ان أي نقص في البيانات سيؤدي الى نقص في المعلومات قيد البحث وبالتالي التأثير على نتائج التحليل المعتمد وخصوصاً أذا كان عدد المشاهدات الاجمالي صغيراً أو عدد المشاهدات الشاذة كبيرًا أو مقارباً الى نصف عدد المشاهدات الكلي.

لهذا السبب فان من الافضل تقدير تلك القيمة اوالقيم الشاذة , وعملية تقدير المشاهدات الشاذة في حالة نماذج الانحدار اللاخطية تحتاج الى مزيد من البحث لأجل تثبيت أداءها ومقدراتها على تقدير المشاهدات الشاذة بشكل جيد .

2. هدف البحث :

يهدف البحث الى دراسة القيم الشاذة في بعض النماذج اللاخطية وتأثير وجودها على مقدرات تلك النماذج ، حيث أن الدراسة تضمنت أكتشاف وتقدير القيمة ألشاذة في حالة المتغير الواحد باقتراح بعض الطرق اللامعلمية.

3. الجانب النظري:

يتضمن هذا المبحث دراسة بعض النماذج اللاخطية فضلاً عن توضيح الصيغ المعتمدة في تحديد التقديرات لمعلمات تلك ال نماذج . وسيتم التطرق ايضاً الى بعض الطرائق اللامعلمية لاكتشاف القيم الشاذة مع اقتراح طرق لتقدير القيم الشاذة.

3-1 نماذج الانحدار اللاخطية : [3]

أن دراسة نماذج الانحدار اللاخطية جاءت لتحديد العلاقة غير الخطية بين متغير الاستجابة (Y) والمعلمات المجهولة (θ).

بافتراض الانموذج ياخذ العلاقة الاتية:

$$Y_i = f(X_i, \theta) + \in_i \qquad i = 1,2,.....,n \quad .... (1)$$

اذ أن

Y_i : يمثل متجه قيم الاستجابة

f : يمثل دالة من Xi و θ أي أن

$$X_i = (x_{i1},........, x_{ik})^T$$

$$\theta = (\theta_1,.............., \theta_P)^T$$

$\in_i$: يمثل الخطأ العشوائي الذي يتوزع طبيعياً بمتوسط صفر وتباين ثابت σ^2

وباستخدام طريقة المربعات الصغرى اللاخطية يمكن تصغير المقدار الاتي

$$S(\theta) = \sum_{i=1}^{n} [y_i - f(X_i, \theta)]^2 \quad .... (2)$$

والحصول على تقديرات المعلمات

$$\hat{\theta} = \theta_o + (F'F)^{-1} F' \in$$

اذ أن

θ_o : يمثل متجه ذات بعد p للمعلمات غير المعلومة

F : يمثل مصفوفة ذات بعد nxp عناصرها عبارة عن المشتقات الجزئية $\dfrac{\partial f(X_i, \theta_o)}{\partial \theta_j}$

$\in$: يمثل متجه البواقي يتوزع طبيعياً بمتوسط صفر وتباين σ^2 وان تقدير تباين الخطأ يكون

$$S^2 = \frac{\in' \left[I - F(F'F)^{-1} F' \right] \in}{n-p}$$

أن توزيع مقدرات المعلمات θ يقترب من التوزيع الطبيعي المتعدد ذات البعد **P** أي أن

$$\hat{\theta} \sim Np\left(\theta_o, \sigma^2 (F'F)^{-1}\right)$$

$$\frac{(n-p)S^2}{\sigma^2} \sim \chi^2_{(n-p)}$$

3-1-1 أنموذج الانحدار اللاخطي في حالة وجود معلمة واحدة غير خطية: [5]

عند ظهور معلمة واحدة بشكل لاخطي يمكن التعبير عن العلاقة اللاخطية بين المتغير التابع (Y_t) والمعلمة غير المعلومة (θ) بالصيغة الآتية:

$$Y_t = X_t^{\theta} + \in_t \qquad t = 1,2,........,n \quad .... (3)$$

اذ أن الخطأ العشوائي $\in_t$ يتوزع طبيعي بوسط صفر وتباين معلوم σ^2 لكل $t = 1,2,.....,n$ وان θ تمثل معلمة الانموذج المراد تقديرها بطريقة المربعات الصغرى (L.S.) وذلك من خلال تصغير المقدار الاتي:

$$S(\theta) = \sum (Y_t - X_t^{\theta})^2 \; .... (4)$$

باشتقاق الصيغة (4) ومساواة المشتقة بالصفر نحصل على تقدير المعلمة θ

$$\frac{\partial S}{\partial \theta} = -2 \sum (Y_t - X_t^{\theta})(\log X_t) X_t^{\theta} = 0$$

$$\sum Y_t \log X_t (X_t^{\hat{\theta}}) = \sum (\log X_t) X_t^{2\hat{\theta}} \quad .... (5)$$

وان القيمة التقديرية الاولية للمعلمة $\hat{\theta}$ تعتمد في اسلوب كاوس نيوتن (Gauss-Newton Algorithms) ضمن برنامج الانحدار اللاخطي Non Linear Regression وبعد عدة دورات يمكن الحصول على مقدر المعلمة $\hat{\theta}$ وباقل مجموع مربعات خطأ. لكن المقدر $\hat{\theta}$ يعتبر مقدر متحيز ولا يمتلك اقل تباين وبذلك لا يمثل توفيق خطي للمتغير Y_t أي بمعنى صفات المقدر

$\hat{\theta}$ غير مطابقة لصفات المقدر $\hat{\beta}$ لحالة الانموذج الخطي $Y_t = \beta X_t + \in_t$ ولكن يمكن اقتراب من خصائص مقدر $\hat{\beta}$ عند زيادة حجم العينة. وان توزيع مقدر $\hat{\theta}$ يقترب من التوزيع الطبيعي.

وكما هو معلوم عند افتراض الاخطاء العشوائية $\in_t$ تتوزع بشكل طبيعي متماثل ومستقل (iidN)، ومع زيادة حجم العينة فان مقدرات المربعات الصغرى (L.S.) تساوي مقدرات طريقة الامكان الاعظم (M.L.) أي الحصول على مقدرات غير متحيزة وتمتلك اقل تباين والتي تسمى (Minimum Variance Bound).

3-1-2 أنموذج الانحدار اللاخطي في حالة وجود معلمتين غير خطية:[4]

عند ظهور معلمتين غير خطية يمكن التعبير عن العلاقة من خلال دالة النمو الاسية البدائية (Initially Grows Exponentially Function) بالمعلمتين λ, γ حسب الزمن t وكما مبين بالصيغة الاتية

$$Y = \gamma\, e^{\lambda t} + \in \qquad \gamma > 0, \lambda > 0 \qquad .... \quad (6)$$

اذ أن

γ : عبارة عن ثابت يتحدد بواسطة بعض الشروط الاولية او انها تمثل القيمة الحدية لحجم المجتمع عند بدء المشاهدة.

λ : تمثل نسبة معدل النمو خلال الزمن t ($\lambda > 1$) .

$\in$: يمثل الخطأ العشوائي يتوزع طبيعي بوسط صفر وتباين معلوم σ^2

ولتقدير معلمتي الانموذج يمكن استخدام طريقة المربعات الصغرى (L.S.) وذلك من خلال تصغير المقدار الآتي

$$S(\theta) = \sum_{i=1}^{n} (y_i - \gamma\, e^{\lambda t_i})^2 \; .... \; (7)$$

3-1-3 نماذج الانحدار اللاخطية في حالة وجود اكثر من معلمتين:

في دراسات منحنيات النمو تظهر عادة اكثر من معلمتين غير خطيتين في الانموذج لذلك فعند تحليل مجموعة بيانات خاصة بالنمو سواء كان في المجال البايولوجي، الزراعي، الهندسي والاقتصادي فان الشكل البياني لمنحنى البيانات يأخذ شكل حرف S أي البدء من نقطة ثابتة ثم تزايد النمو تدريجياً والوصول الى نقطة ألتواء ثم الانخفاض عند النقطة الاخيرة، وتم استخدام مجموعة

دوال رياضية للتعبير عن تلك النماذج ذات الشكل حرف S ومن تلك النماذج تم دراسة أنموذج Gompertz وأنموذج Logistic وأنموذج Weibull وهي كما يلي:

1- أنموذج Gompertz :

يكتب بالصيغة الاتية

$$Y = \alpha \exp[-\exp(\beta - \gamma X)] \quad \ldots\ldots\ldots\ldots\ldots\ (8)$$

2- أنموذج Logistic:

يكنب بالصيغة الاتية

$$Y = \frac{\alpha}{1 + \exp(\beta - \gamma X)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ (9)$$

3- أنموذج Weibull:

يكتب بالصيغة الاتية

$$Y = \alpha - \beta \exp(-\gamma X^{\delta}) \quad \ldots\ldots\ldots\ldots\ldots\ (10)$$

فضلاً عن توضيح كيفية تحديد سلسلتين من التقديرات للمعلمات رغم اختلافات الفرضيات.

الاولى تمثل تقديرات المعلمات مع فرضية حالة الجمع بالنسبة لحد الخطأ لتلك النماذج والتي يعبر عنها كالاتي

$$Y_{tA} = f(X_t, \underline{\theta}) + \in_{tA} \quad \ldots\ldots\ldots\ldots\ldots\ (11)$$

اذ أن $\underline{\theta}$ تمثل متجه المعلمات γ, β, α (عندما δ مناسبة) والتي يتم تقديرها عندما $\in_{tA}$ يتبع التوزيع الطبيعي المتماثل(iidN)بمتوسط صفر وتباين (غير معلوم) σ_A^2.

الثانية تمثل تقديرات المعلمات مع فرضية حالة الضرب بالنسبة لحد الخطأ أي باستخدام التحويل اللوغاريتمي للنماذج ويمكن التعبير عنها كالآتي

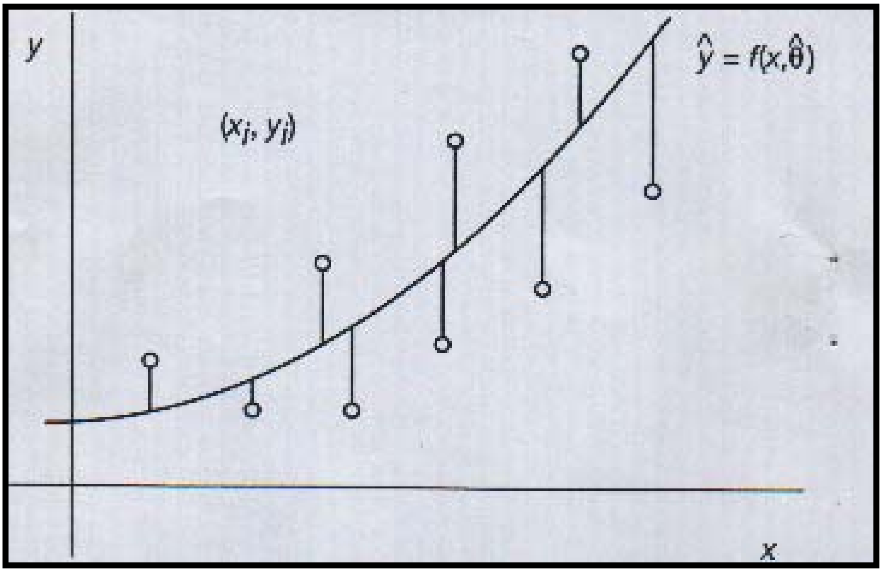
$$Log\, Y_{tM} = \log f(X_t, \underline{\theta}) + \in_{tM} \quad \ldots\ldots\ldots\ (12)$$

حيث $\in_{tM}$: يتبع التوزيع الطبيعي المتماثل (iidN) .بمتوسط صفر وتباين (غير معلوم) σ_M^2 .

3-2 تمثيل مجال المتغير والمشاهدة :

Observation and Variable Space Representations

الشكل رقم (1) يوضح مجال المشاهدة من حيث مشكلة تصغير المقدار $S(\theta)$ حسب الصيغة (2)، وان أسلوب اختيار قيمة $\hat{\theta}$ التي بدورها تساعد في تصغير مجموع مربعات الخط العمودي الساقط على المنحنى، حسب نقاط البيانات المشاهدة (x_i , y_i) .



شكل (1)

يوضح مجال المشاهدة في مشكلة تقدير المربعات الصغرى اللاخطية

لتوضيح اكثر يمكن كتابة أنموذج (1) في صيغة متجه كالآتي:

$y = f(\theta) + e$ .... (13)

حيث كل من $f(\theta), y$ تمثل متجهات ضمن المجال R^n ولكل مركبة ith يوجد y_i ، $f_i(\theta) = f(X_i , \theta)$ و $\in_i$ على التوالي.

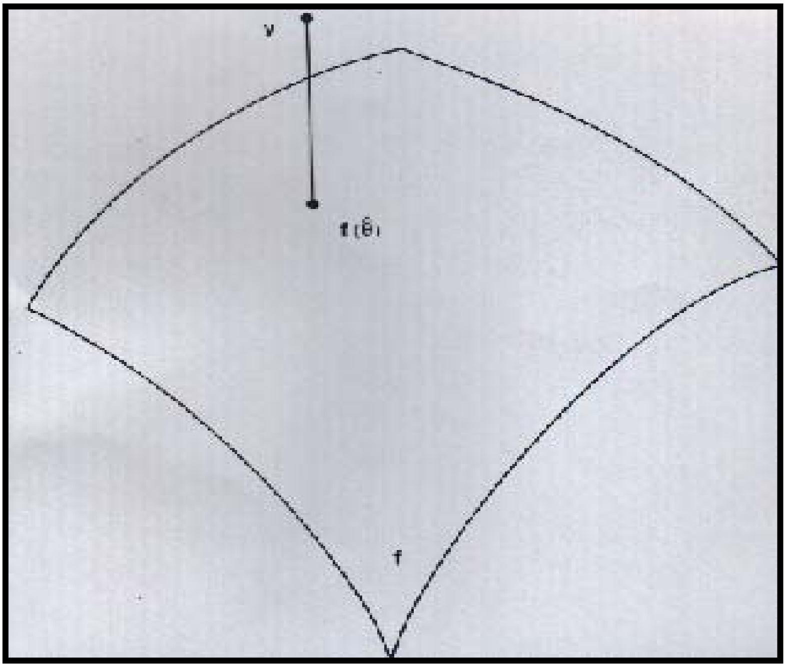
بهذا يمكن التعبير عن الصيغة (1) كما يلي

$Q(\theta) = \|y - f(\theta)\|^2 \qquad \theta \in \Theta$ .... (14)

لكل θ في المجال Θ .

أن مشكلة المربعات الصغرى هو ايجاد تقدير $\hat{\theta}$ حيث يجعل تقارب $f(\hat{\theta})$ الى متجه y (متغير الاستجابة المشاهد)، والشكل رقم (2) يوضح مجال المتغير المقدر بطريقة المربعات الصغرى.

بما أن Θ يمثل مجال المعلمة θ ومجال $f(\theta)$ ذات بعد p في R^n . فإن افضل مقدر لـ $\hat{\theta}$ يحقق حالة التقارب حيث يكون $\hat{y} = f(\hat{\theta})$ أي تاتي نقطة مطابقة الى y .



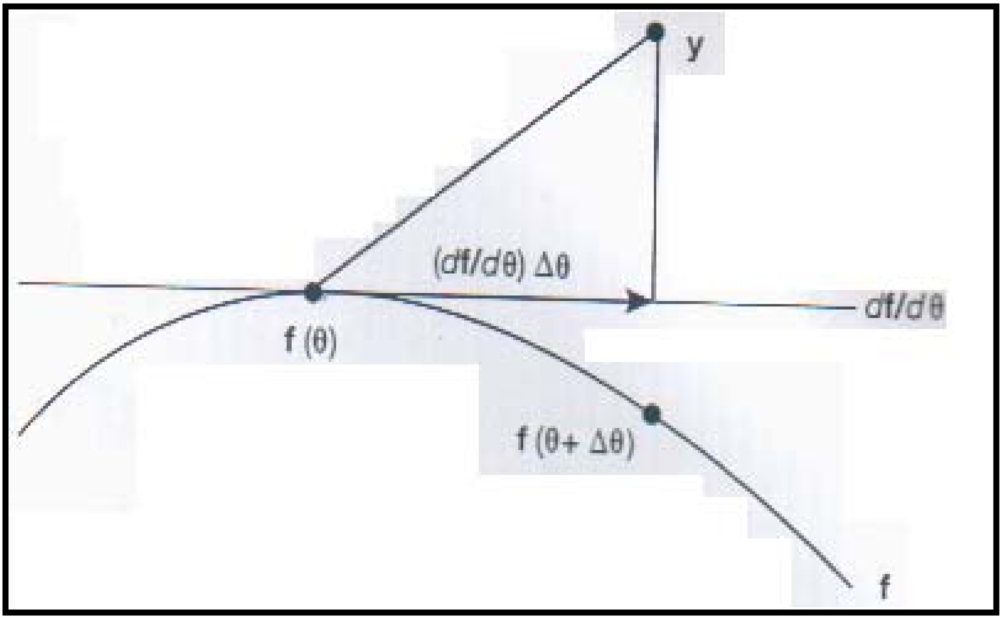
شكل (2)

يوضح مجال المتغير في مشكلة تقدير المربعات الصغرى اللاخطية

3-2 طريقة كاوس نيوتن: The Gauss –Newton Algorithm

تعد طريقة كاوس نيوتن الافضل من بين الطرق الاخرى في تصغير مقدار مربع انحرافات قيم المشاهدات . لذلك تستخدم هذه الطريقة في الحصول على مقدرات المربعات الصغرى، وتعد مقياس دقيق ومناسب لكل من السببين الآتيين .

الاول لانـه يـشترط مجـال طبيعـي فـي الجانـب النظـري للانحـدار الخطـي والانحـدار اللاخطـي والثاني لان اغلب طرق الانحدار اللاخطية تعتمد صيغ معدلة لطريقة كاوس نيوتن فضلاً عن بساطة وسهولة الطريقة في حل مشاكل عدة. والشكل رقم (3) يوضح فكرة اسلوب كاوس نيوتن.



شكل (3)

يوضح مجال المتغير في اسلوب كاوس نيوتن

اذ ان f يمثل مجال متجهات الاستجابة .نبدأ بنقطة ثابتة θ مع ثبوت النقطة $f(\theta)$ على مجال الدالـة. وان مستوي المماس موضح في النقطة $f(\theta)$ من خلال $\frac{df}{d\theta}$ لكون اعمدة $\frac{df}{d\theta}$ تمثل امتداد الى مستوي المماس.

يمكن تمثيل الطريقة من خلال $y - f(\theta)$ على المستوي المماس للمتجه $\left(\frac{df}{d\theta}\right)\Delta\theta$ لتصغير المقدار الاتي .

$$Q_\theta(\Delta\theta) = \left\| y - f(\theta) - \frac{df}{d\theta}\Delta\theta \right\|^2 \quad .... (15)$$

بافتراض اعمدة $\frac{df}{d\theta}$ مستقلة خطياً وان

$$\Delta\theta = \left[\frac{df^T}{d\theta}\cdot\frac{df}{d\theta}\right]^{-1} \frac{df^T}{d\theta}(y - f(\theta)) \quad .... (16)$$

وعمليـة التعويض المتعاقب تبدأ من خـلال اسـتبدال θ بالحد $\theta + \Delta\theta$ للحصول علـى نقطة جديدة $f(\theta + \Delta\theta)$ واقعة على مجال f .

وحسب تعريف الدالة $Q_\theta(\Delta\theta)$ يمكن حل مشكلة التقدير بطريقـة المربعات الصغرى في كل خطوة والحصول على تقارب $Q(\theta + \Delta\theta)$ وذلك باستبدال $f(\theta + \Delta\theta)$ في الصيغة (14) الى $Q(\theta + \Delta\theta)$ وحسب متسلسلة تايلر يكون

$$f(\theta + \Delta\theta) \cong f(\theta) + \frac{df}{d\theta}\Delta\theta$$

4-3 تحديد التقديرات الاولية المناسبة للمعلمات:

Obtaining Good Initial Parameter Estimates

تـم دراسـة موضـوع تقدير المعلمـات فـي النمـاذج اللاخطيـة مـن قبل **Mead** عـام **(1970)** وذلـك باسـتخدام طريقـة المربعـات الـصغرى المرجحـة كتقريـب لطريقـة المربعـات الـصغرى الاعتياديـة وباعتماد طريقة المربعات الصغرى المرجحة في تحديد التقديرات الاولية عند اختبار فرضية حد الخطأ لحالة الجمع عندما $Var(Y_t) = \sigma^2$ وايضاً لتصغير قيمة المقدار الآتي:

$$S = \sum_{t=1}^{n} \frac{\sigma^2}{Var(Y_t)}\left[Y_t - E(Y_t)\right]^2 \quad .... (17)$$

امـا عنـد اختبـار الفرضـية بالنـسبة لحالـة الـضرب حيـث $Var(\log Y_t) = \sigma^2$ لكـل **t=1,2,……,n** وباستبدال $Var(Y_t)$ في الصيغة **(17)** بالمقدار الاتي:

$$Var\left[f(Y_t)\right] \simeq \left[\frac{\partial f(Y)}{\partial Y_t}\right]^2_{Y_t = E(Y_t)} \cdot Var(Y_t) \quad .... (18)$$

فإن ناتج الاشتقاق يكون بدلالة $E(Y_t)$ عندما $Y_t = E(Y_t)$ ومن الصيغة **(18)** يمكن الحصول على $Var(Y_t)$ وكما يلي

$$\operatorname{var}(Y_t) \cong \frac{Var[f(Y_t)]}{[\partial f(Y_t)/\partial Y_t]^2_{Y_t=E(Y_t)}} \quad .... (19)$$

عندما $f(Y_t) = \log Y_t$ فإن الصيغة (19) تكون

$$\operatorname{var}(Y_t) = \frac{\operatorname{var}(Log Y_t)}{\left(\frac{1}{Y_t}\right)^2_{Y_t=E(Y_t)}} \quad .... (20)$$

اذ أن

$$\operatorname{var}(Y_t) = \sigma^2 E^2(Y_t) \, ... (21)$$

وبتعويض var (Yt) حسب الصيغة (21) في الصيغة (17) نحصل على

$$S = \sum_{t=1}^{n} \left[\frac{Y_t}{E(Y_t)} - 1\right]^2 \, ... (22)$$

وفي عام (1974) تطرق Bard بالتفصيل الى عملية تقدير المعلمات في النماذج اللاخطية وكذلك كل منKennedy , Gentle في عام (1980) . من تلك الدراسات تبين لنا أن الدقة مطلوبة عند تحديد التقديرات الاولية للمعلمات عند دراسة النماذج اللاخطية لتمثيل مجموعة من البيانات ، من اجل تحقيق الشروط المثلى لمقدرات المربعات الصغرى للمعلمات ، غير أن الاختيار غير الملائم للقيم الاولية يجعل تحقيق تلك الشروط امراً صعب .

وفيما يلي نوضح أشتقاق الصيغ المعتمدة في تحديد التقديرات الاولية لأنموذج الأتي.

3-4-1 أنموذج الانحدار التقريبي[5] Asymptotic Regression Model

يأخذ الصيغة الاتية :

$$Y = \alpha - \beta \gamma^X \qquad ...(23)$$

وعندما نأخذ اللوغارتم الطبيعي للصيغة اعلاه نحصل على معادلة انحدار Z على X الاتية:

$$Z = Log(\alpha_0 - Y) = Log\,\beta + X\,Log\,\gamma \qquad ...(24)$$

بنفس الاسلوب المستخدم في الانموذج السابق يمكن تحديد التقدير الاولي للمعلمة α_0

ومن حل معادلة انحدار (24) يمكن الحصول على التقديرات الاولية للمعلمتين $Log\,\beta$ و $Log\,\gamma$ وكما يلي :

$$Log\,\gamma_0 = \frac{n\sum X_i\, Z_i - \sum X_i \sum Z_i}{n\sum X_i^{\,2} - \left(\sum X_i\right)^2}$$

$$\gamma_0 = \exp\ \frac{n\sum X_i\, Z_i - \sum X_i \sum Z_i}{n\sum X_i^{\,2} - \left(\sum X_i\right)^2} \qquad \text{...(25)}$$

$$\log \beta_0 = \bar{Z} - (Log\,\gamma_0)\,\bar{X}$$

او

$$\beta_0 = \exp\left(\bar{Z} - (Log\,\gamma_0)\,\bar{X}\right) \qquad \text{...(26)}$$

وباستخدام القيم الاولية للمعلمات $\gamma_0, \beta_0, \alpha_0$ في اولى خطوات كاوس نيوتن يمكن الحصول على افضل المقدرات وبأقل مجموع مربعات للبواقي اي ان :

$$RSS = \sum_{t=1}^{n}\left(Y_t - \alpha_0 + \beta_0 \gamma_0^{X_t}\right)^2 \qquad \text{...(27)}$$

في حالة اختبار الفرضية لحالة الجمع يكون:

$$RSS = \sum_{t=1}^{n}\left(Log\,Y_t - Log(\alpha_0 - \beta_0 \gamma_0^{X_t})\right)^2 \qquad \text{...(28)}$$

في حالة اختبار الفرضية لحالة الضرب.

وان من الممكن تحديد التقديرات الاولية للمعلمات بطريقة اخرى وذلك بتعويض قيمة Y حسب انموذج (23) في الصيغة (22) يكون :

$$S = \sum_{t=1}^{n}\left(Y - \alpha + \beta \gamma^{X}\right)^2 \qquad \text{...(29)}$$

وباشتقاق الصيغة اعلاه للمعلمتين α و β ومساواة المشتقة بالصفر نحصل على المعادلتين الاتيتين:

$$\begin{aligned} n\alpha_0 - \beta_0 \sum \gamma_o^{X} &= \sum Y \\ \alpha_0 \sum \gamma_o^{X} - \beta_0 \sum \gamma_o^{2X} &= \sum \gamma_0^{X} Y \end{aligned} \qquad \text{...(30)}$$

يتم تحديد التقدير الاولي للمعلمة γ ويرمز لها γ_0 بحيث تتراوح بين الصفر والواحد وثم من حل المعادلتين الخطيتين (30) يمكن الحصول على التقديرات الاولية للمعلمتين α و β وكما يلي:

$$\beta_0 = \frac{n\sum \gamma_0^X Y - \sum Y \sum \gamma_0^X}{\left(\sum \gamma_0^X\right)^2 - n\sum \gamma_0^{2X}} \qquad ...(31)$$

$$\alpha_0 = \frac{\sum Y + \beta_0 \sum \gamma_0^X}{n} \qquad ....(32)$$

ثم القيم الاولية للمعادلات $\gamma_0 , \beta_0 , \alpha_0$ سوف يعتمد في اولى خطوات كاوس نيوتن للتوصل الى تقديرات المعادلات $\hat{\gamma} , \hat{\beta} , \hat{\alpha}$

اضافة لما ورد اعلاه لقد تم دراسة نماذج الانحدار اللاخطية في حالة وجود اكثر من معلمتين مثلا انموذج Richards وفي هذا االبحث تم استخدام انموذج الانحدار التقريبي مع تطبيق عملي نظراً لاهمية النماذج غير الخطية من ناحية ولقلة الدراسات على النماذج اللاخطية مع تطبيقات عملية .

3-5 طرق اكتشاف وتقدير القيم الشاذة

3-5-1 الطرق اللامعلمية (طريقة Tukey) لاكتشاف القيم الشاذة:[6]

وهي تلك الطرائق التي لاتفترض توزيع معين على البيانات متخطية بذلك احد الافتراضات الاساسية التي تعتمد عليها الطرائق الاخرى (المعلمية) ومن اهم هذه الطرائق هي طريقة الصندوق والقطع المخططة مع الملخصات الخمسة (The Box and Whisker Plots with 5 Number summaries) المقترحة من قبل Tukey عام 1977 والتي تستخدم لفحص القيمة او القيم الشاذة في حالة المتغير الواحد وبافتراض عدم معرفة التوزيع الاحتمالي للمتغير ويمكن تلخيص هذه الطريقة كالاتي:

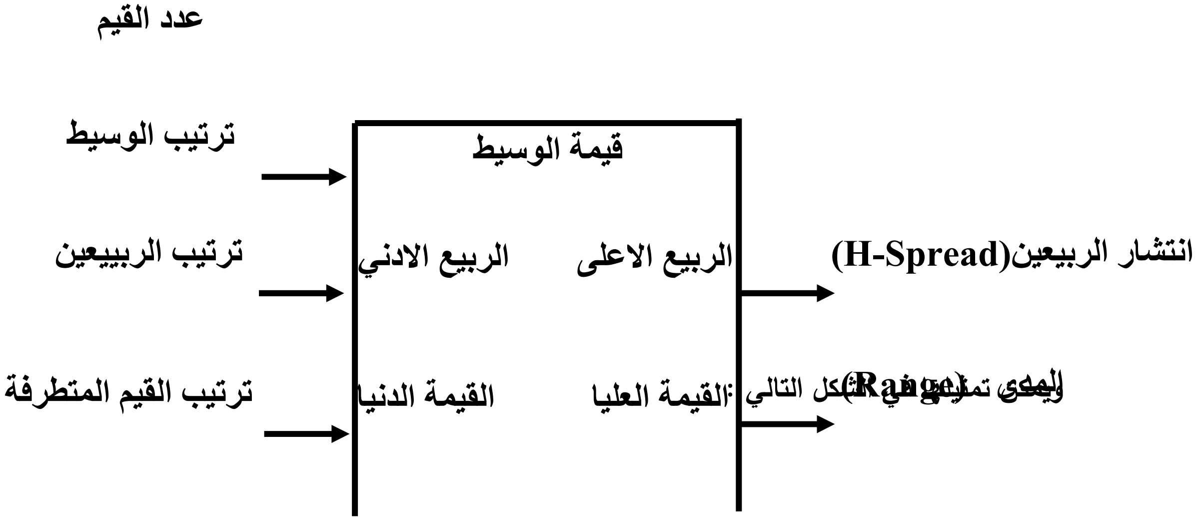
1. الملخصات الخمسة Five Number Summaries

تشمل على خمسة عناصر اساسية وهي:

- قيمة الوسيط μ_e The Median

- **قيمة الربيع الاعلى Q_3 Upper Hinge**
- **قيمة الربيع الادنى Q_1 Lower Hinge**
- **قيمة عليا متطرفة Upper Extreme**
- **قيمة دنيا متطرفة Lower Extreme**

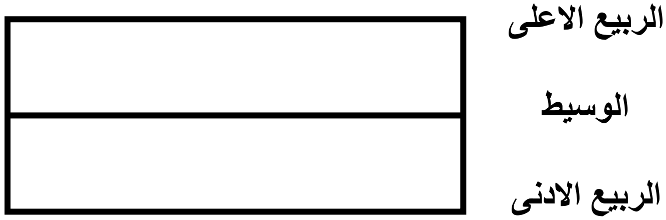
ويمكن تمثيلها في الشكل التالي:



شكل رقم (4)
(يوضح مخطط الملخصات الخمسة)

2. الصندوق والقطع المخططة (Box and Whisker Plots):

لتمثيل البيانات من خلال اطار معين (Box) ، يحده من الاعلى الربيع الاعلى ومن الاسفل الربيع الادنى وان والوسيط Median يكون موقعه بين الربيعين كما في الشكل رقم (5).



شكل رقم (5)

(يوضح الصندوق والقطع المخططة)

اما القطع المخططة (Whisker) تمثل حدود اخرى غير الصندوق (Box) لتوضيح انتشار البيانات وترسم من نهاية الصندوق لتحديد السياجين الداخلي والخارجي لكل من الطرفين وتستخدم في تمثيل التطرف في القيم ولتطبيق الطريقة يمكن تلخيص الخطوات كما يلي:

1. ترتيب قيم المتغير المراد دراسته تصاعدياً.
2. تحديد رتبة الوسيط والمساوية الى $(n+1)/2$ ثم تحديد قيمة الوسيط والتي هي القيمة التي تتوسط البيانات اذا كان العدد فردياً ويمثل متوسط القيمتين الوسطيتين اذا كان العدد زوجياً.
3. تحديد الربيع الادنى Q_1 (Lower Hinge) هو القيمة التي ترتيبها $\frac{n}{4}$ من بين القيم المرتبة تصاعدياً ، وقيمة الربيع الاعلى Q_3 (Upper Hinge) هو القيمة التي ترتيبها $\frac{3n}{4}$ من بين القيم المرتبة تصاعدياً.
4. تحديد قيمة انتشار الربيعين (H-Spread) هو القيمة الناتجة من الفرق بين قيمتي الربيعين حيث:

$$(H-Spread)=Q_3-Q_1$$

5. تحديد قيمة الخطوة (Step) والتي هي عبارة عن انتشار الربيعين مضروباً في (1.5) حيث

$$\begin{aligned} Step &= 1.5\,(H-Spread) \\ &= 1.5\,(Q_3-Q_1) \end{aligned}$$

6. تحديد السياج الداخلي (Inner Fence) ويمثل بعد خطوة واحدة ابتداءاً من الربيع الاعلى وصعوداً او الربيع الادنى ونزولاً.

(السياج الداخلي من الاعلى) $= Q_3 + Step$

(السياج الداخلي من الاسفل) $= Q_1 - Step$

7. تحديد السياج الخارجي (Outer Fence) ويمثل بعد خطوتين ابتداءاً من الربيع الاعلى او الادنى من الجهة التالية:

$$Q_3 + 2(Step) = \text{(السياج الخارجي من الاعلى)}$$

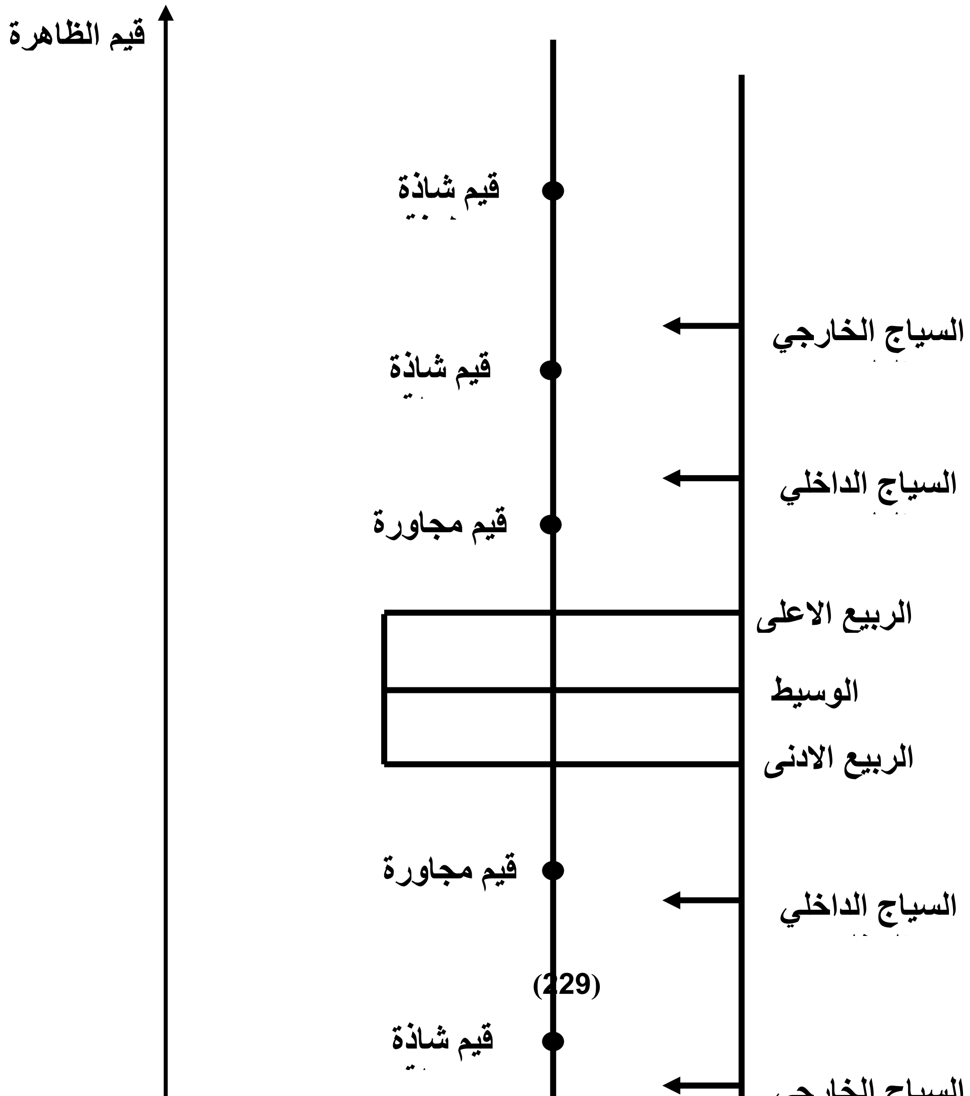
$$Q_1 - 2(Step) = \text{(السياج الخارجي من الاسفل)}$$

8. القيم المجاورة (Adjacent Values) : هي القيم القريبة من السياج الداخلي العلوي (السفلي) ولكنها لازالت داخلة.

9. قيم شاذة معتدلة (Mild Outliers) : هي القيم الواقعة بين السياج الداخلي العلوي (السفلي) والسياج الخارجي العلوي (السفلي).

10. قيم شاذة متطرفة (Extreme Outliers) : هي القيم الواقعة خارج حدود السياج الخارجي.

والشكل رقم (6) يوضح مخطط الصندوق والقطع المخططة مع الملخصات الخمسة لتحديد القيم الشاذة.



شكل رقم (6)

يوضح مخطط الصندوق والقطع المخططة مع الملخصات الخمسة

3-5-2 الطرق المقترحة لتقدير القيم الشاذة :

في هذا المبحث تم اقتراح طرق جديدة لتقدير القيمة الشاذة في حالة المتغير الواحد وذلك من خلال استخدام مقاييس تمثيل البيانات والتي تتأثر بوجود القيمة الشاذة هما المتوسط والانحراف المتوسط اللذان يعطيان فكرة عن تمركز القيم وانتشارها ويمكن توضيح الطرق المقترحة للتقدير كما يلي:

a. الطريقة المقترحة الاولى :

يتم التعويض بدل القيمة الشاذة في الكبر بقيمة جديدة تساوي قيمة انتشار الربيعين + قيمة الوسيط – 1 وفي حالة وجود اكثر من قيمة في هذا الطرف من القيم فتطرح من القيمة (انتشار الربيعين + الوسيط) 1 ، 2 وهكذا. اما اذا كانت القيمة الشاذة هي في الطرف الاصغر فيتم اضافة واحد الى الفرق بين الوسيط وقيمة انتشار الربيعين في حالة وجود قيمة شاذة واحدة ونضيف 1 ، 2 ، 3 وهكذا في حالة وجود اكثر من قيمة واحدة اي ان:

$$d_i^* = H + \mu_e - 1 \qquad ...(33)$$

اذا d_i شاذة في الكبر

$$d_i^* = \mu_e - H + 1 \qquad ...(34)$$

اذا d_i شاذة في الصغر

اذ ان H : يمثل قيمة انتشار الربيعين ؛ μ_e : يمثل قيمة الوسيط

b. الطريقة المقترحة الثانية :

يتم التعويض بدل القيمة الشاذة في الكبر بقيمة جديدة تساوي قيمة انتشار الربيعين + قيمة الانحراف المتوسط (بالوسط الحسابي) −1 وفي حالة وجود اكثر من قيمة في هذا الطرف من القيم فتطرح من القيمة (انتشار الربيعين) + الانحراف المتوسط (بالوسط الحسابي)) 1 ، 2 وهكذا. اما اذا كانت القيمة الشاذة هي من الطرف الاصغر فيتم اضافة واحد الى الفرق بين الانحراف المتوسط وقيمة انتشار الربيعين في حالة وجود قيمة شاذة واحدة ونضيف 2 ، 3 وهكذا في حالة وجود اكثر من قيمة واحدة اي ان:

$$d_i^* = H + M.D(\bar{Y}) - 1 \qquad ...(35)$$

اذا d_i شاذة في الكبر

$$d_i^* = M.D(\bar{Y}) - H + 1 \qquad ...(36)$$

اذا d_i شاذة في الصغر

اذ ان $M.D(\bar{Y})$ يمثل الانحراف المتوسط (بالوسط الحسابي) اي ان

$$M.D(\bar{Y}) = \frac{\sum_{i=1}^{n} \left| Y_i - \bar{Y} \right|}{n}$$

c. الطريقة المقترحة الثالثة :

يتم التعويض بدل القيمة الشاذة في الكبر بقيمة جديدة تساوي قيمة انتشار الربيعين + قيمة الانحراف المتوسط (بالوسيط) −1 وفي حالة وجود اكثر من قيمة في هذا الطرف من القيم فنطرح من القيمة (انتشار الربيعين + الانحراف المتوسط (بالوسيط)) 1 ، 2 وهكذا. اما اذا كانت

القيمة الشاذة هي في الطرف الاصغر فيتم اضافة واحد الى الفرق بين الانحراف المتوسط وقيمة انتشار الربيعين في حالة وجود قيمة شاذة واحدة ونضيف 2 ، 3 وهكذا في حالة وجود اكثر من قيمة واحدة اي ان :

$$d_i^* = H + M.D(\mu_e) - 1 \qquad \ldots(37)$$

اذا d_i **شاذة في الكبر**

$$d_i^* = M.D(\mu_e) - H + 1 \qquad \ldots(38)$$

اذا d_i **شاذة في الصغر**

اذ ان $M.D(\mu_e)$ **يمثل الانحراف المتوسط (بالوسيط)**

$$M.D(\mu_e) = \frac{\sum_{i=1}^{n} \left| Y_i - \mu_e \right|}{n}$$

d. <u>الطريقةالمقترحة الرابعة :</u>

يتم التعويض بدل القيمة الشاذة في الكبر بقيمة جديدة تساوي قيمة انتشار الربيعين + (الربيع الاعلى – الوسيط) −1 وفي حالة وجود اكثر من قيمة في هذا الطرف من القيم فنطرح من القيمة (انتشار الربيعين + (الربيع الاعلى – الوسيط)) 1 ، 2 وهكذا. اما اذا كانت القيمة الشاذة هي في الطرف الاصغر فيتم اضافة واحد الى الفرق بين (الربيع الاعلى – الوسيط) وقيمة انتشار الربيعين في حالة وجود قيمة شاذة واحدة ونضيف 2 ، 3 وهكذا في حالة وجود اكثر من قيمة واحدة اي ان :

$$d_i^* = H + \left(Q_3 - \mu_e\right) - 1 \qquad \ldots(39)$$

اذا d_i **شاذة في الكبر**

$$d_i^* = \left(Q_3 - \mu_e\right) - H + 1 \qquad \ldots(40)$$

اذا d_i **شاذة في الصغر**

e. <u>الطريقة المقترحة الخامسة :</u>

يتم التعويض بدل القيمة الشاذة في الكبر بقيمة جديدة تساوي قيمة انتشار الربيعين + الوسط الحسابي −1 وفي حالة وجود اكثر من قيمة في هذا الطرف من القيم فنطرح من القيمة (انتشار الربيعين + الوسط الحسابي) 1 ، 2 وهكذا. اما اذا كانت القيمة الشاذة هي في الطرف

الاصغر فيتم اضافة واحد الى الفرق بين الوسط الحسابي وقيمة انتشار الربيعين في حالة وجود قيمة شاذة واحدة ونضيف 2 ، 3 وهكذا في حالة وجود اكثر من قيمة واحدة اي ان:

$$d_i^* = H + \bar{Y} - 1 \qquad ...(41)$$

اذا d_i شاذة في الكبر

$$d_i^* = \bar{Y} - H + 1 \qquad ...(42)$$

اذا d_i شاذة في الصغر

f. الطريقة المقترحة السادسة:

يتم التعويض بدل القيمة الشاذة في الكبر بقيمة جديدة تساوي قيمة انتشار الربيعين + (الربيع الاعلى – الخطوة) -1 وفي حالة وجود اكثر من قيمة في هذا الطرف من القيم فنطرح من القيمة (انتشار الربيعين + (الربيع الاعلى – الخطوة)) 1 ، 2 وهكذا. اما اذا كانت القيمة الشاذة هي الطرف الاصغر فيتم اضافة واحد الى الفرق بين (الربيع الاعلى-الخطوة) وقيمة انتشار الربيعين في حالة وجود قيمة شاذة واحدة ونضيف 2 و 3 وهكذا في حالة وجود اكثر من قيمة واحدة اي ان :

$$d_i^* = H + \left(Q_3 - S\right) - 1 \qquad ...(43)$$

اذا d_i شاذة في الكبر

$$d_i^* = \left(Q_3 - S\right) - H + 1 \qquad ...(44)$$

اذا d_i شاذة في الصغر

g. الطريقة المقترحة السابعة :

يتم التعويض بدل القيمة الشاذة في الكبر بقيمة جديدة تساوي قيمة انتشار الربيعين + (الخطوة-الوسيط) -1 وفي حالة وجود اكثر من قيمة في هذا الطرف من القيم فنطرح من القيمة (انتشار الربيعين + (الخطوة-الوسيط)) 1 و 2 وهكذا . اما اذا كانت القيمة الشاذة هي في الطرف الاصغر فيتم اضافة واحد الى الفرق بين (الخطوة – الوسيط) وقيمة انتشار الربيعين في حالة وجود قيمة شاذة واحدة ونضيف 2 ، 3 وهكذا في حالة وجود اكثر من قيمة واحدة اي ان :

$$d_i^* = H + \left(S - \mu_e\right) - 1 \qquad ...(45)$$

اذا d_i شاذة في الكبر

$$d_i^* = \left(S - \mu_e\right) - H + 1 \qquad ...(46)$$

اذا d_i شاذة في الصغر.

4. الجانب التطبيقي

في هذا الجانب تم دراسة حالة القيمة غير المنسجمة والظاهرة في البيانات الموضحة في الجدول رقم (1) التي تبين دراسة معدل نمو نبات الحنطة نوع صابربيك خلال فترة زمنية محددة بالايام وهي تجربة أقيمت في البيوت الزجاجية لكلية العلوم –الجامعة المستنصرية .

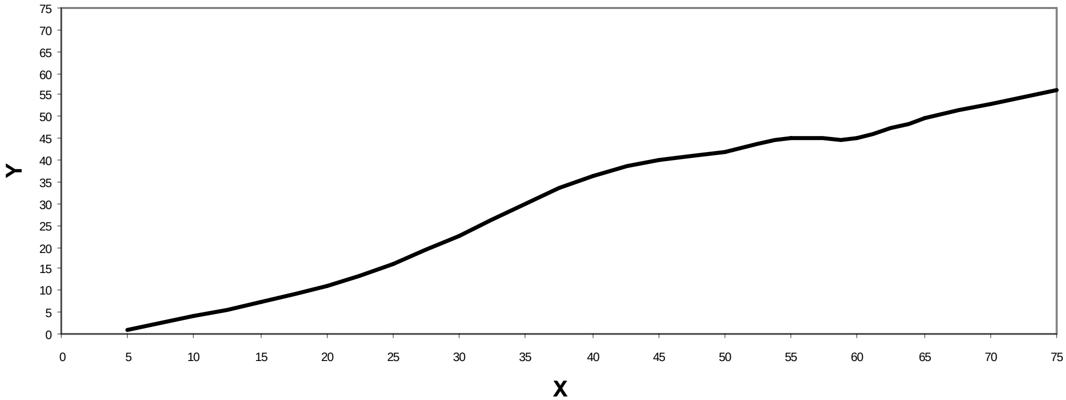
جدول رقم (1)

يوضح بيانات نمو نبات الحنطة حسب الايام

معدل نمو اطول خمسة نباتات (سم) Y	الزمن بالايام X
1	5
4	10
7.5	15
11	20
16	25
22.5	30
30	35
36.5	40
40	45
42	50
45	55
45.3	60
49.5	65
52.9	70
56.2	75

4-1 التحليل الاحصائي :

هنا تم استخدام انموذج الانحدار التقريبي **Asymptotic Regression Model** الذي كثيراً ما يستخدم لتحليل بيانات نمو النباتات في المجال الزراعي ويعتبر مشابه الى حد ما لنماذج النمو التي تم دراستها عند تحليل بيانات التطبيق الاول من حيث شكل المنحنى يبدأ من نقطة ثابتة وثم يتزايد مع زيادة قيم X لكن يختلف في نقطة الالتواء اي في هذه الحالة شكل المنحنى لاياخذ شكل حرف S . وقبل تحليل البيانات تم توضيحها من خلال الشكل رقم (7).



شكل رقم (7)

يوضح حالة النمو لبيانات الحنطة حسب الايام

ولتقدير معلمات انموذج الانحدار التقريبي تم تحديد سلسلتين من التقديرات للمعلمات على اساس فرضية حدا الخطأ لحالتي الجمع والضرب من خلال تحديد التقدير الاولي للمعلمة α_0 بحيث تقترب من اكبر قيمة لـ Y ومن معادلة انحدار Z على X حسب الصيغة (24) يمكن الحصول على التقديرات الاولية لكل من المعلمتين $Log\,\gamma$ و $\log\beta$ والتي يرمز لها γ_0 و β_0 و حسب الصيغتين (25) و (26) على التوالي ، وكانت مساوية الى $\alpha_0 = 57$ ، $\beta_0 = 115.16$ و $\gamma_0 = 0.954$ وبعد تنفيذ برنامج الانحدار اللاخطي في نظام SPSS تم الحصول على مقدرات المربعات الصغرى عند اختبار فرضية حد الخطأ لحالتي الجمع والضرب بالاضافة الى قيمتي MSE و R^2 وهي كما يلي:

Add.

$$\hat{\alpha} = 105.78, \hat{\beta} = 114.21, \hat{\gamma} = 0.988, MSE = 6.3081, R^2 = 0.98$$

Mult.

$$\hat{\alpha} = 93.653, \hat{\beta} = 106.911, \hat{\gamma} = 0.987, MSE = 0.012, R^2 = 0.98$$

ويلاحظ انخفاض قيمة MSE في حالة الضرب عن حالة الجمع.

4-2 التطبيق العملي للطرق المقترحة لتقدير القيم الشاذة:

لقد تم استخدام الطرق المقترحة في تقدير القيم الشاذة في اكثر من تطبيق واحد ولتحليل بيانات مختلفة في سبيل ملاحظة مدى ملائمة تلك الطرائق في تقدير القيمة الشاذة.

تبين من بيانات جدول رقم (1) ان قيم المتغير الاول تمثل فترة زمنية (بالايام) وهي قيم متزايدة بصورة منتظمة ولانتوقع ان يظهر فيها قيمة شاذة. لذا تم فحص قيم المتغير الثاني باعتماد احدى طرق اكتشاف المشاهدات الشاذة للمتغير الواحد وهي طريقة الصندوق والقطع المخططة مع الملخصات الخمسة التي تم توضيحها في المبحث السابق ، وتم التأكد من عدم وجود قيمة شاذة في البيانات.

لغرض توليد قيمة شاذة، لقد استبدلت قيمة Y الرابعة عشرة من 52.9 الى 98.9 والجدول رقم (2) يوضح البيانات مع وجود القيمة الشاذة

جدول رقم (2)

يوضح البيانات مع وجود القيمة الشاذة

X	Y
5	1
10	4
15	7.5
20	11
25	16
30	22.5
35	30
40	36.5
45	40
50	42
55	45
60	45.3
65	49.5
70	98.9
75	56.2

ولتحليل بيانات جدول رقم (2) تم استخدام انموذج الانحدار التقريبي حسب الصيغة (23)، ولتقدير الانموذج تم تحديد المقدرات الاولية للمعلمات بنفس الاسلوب المبين في المبحث السابق وكانت تساوي $\gamma_0 = 0.953, \beta_0 = 260.91, \alpha_0 = 98.91$ وبعد تنفيذ برنامج الانحدار

اللاخطي في نظام SPSS تم الحصول على مقدرات المربعات الصغرى للمعلمات عند فرضية حد الخطأ لحالتي الجمع والضرب كما موضح في الجدول رقم (2). ونلاحظ زيادة في قيمة MSE في حالة الجمع.

بالامكان معالجة القيمة الشاذة في الكبر من خلال تقدير قيمة الانحراف الشاذ بالطرق المقترحة وكما يلي:

a. تقدير القيمة الشاذة بالطريقة الاولى حسب الصيغة (33) يكون :

$$d_i^* = 69.8$$

b. تقدير القيمة الشاذة بالطريقة الثانية حسب الصيغة (35)يكون :

$$d_i^* = 52.5$$

c. تقدير القيمة الشاذة بالطريقة الثالثة حسب الصيغة (37) يكون :

$$d_i^* = 52.3$$

d. تقدير القيمة الشاذة بالطريقة الرابعة حسب الصيغة (39) يكون :

$$d_i^* = 42.1$$

e. تقدير القيمة الشاذة بالطريقة الخامسة حسب الصيغة (41) يكون :

$$d_i^* = 67$$

f. تقدير القيمة الشاذة بالطريقة السادسة حسب الصيغة (43)يكون :

$$d_i^* = 27.15$$

g. تقدير القيمة الشاذة بالطريقة السابعة حسب الصيغة (45) يكون :

$$d_i^* = 48.3$$

بعد تقدير القيمة الشاذة بالطرق المبينة اعلاه، تم استخدام طريقة الصندوق والقطع المخططة مع الملخصات الخمسة للتأكد من عدم وجود قيمة شاذة. والجدول رقم (3) يوضح البيانات بعد تقدير القيمة الشاذة:

جدول رقم (3)

يوضح البيانات بعد تقدير القيمة الشاذة

X	a	b	c	D	e	f	g

1	1	1	1	1	1	1	1
2	4	4	4	4	4	4	4
3	7.5	7.5	7.5	7.5	7.5	7.5	7.5
4	11	11	11	11	11	11	11
5	16	16	16	16	16	16	16
6	22.5	22.5	22.5	22.5	22.5	22.5	22.5
7	30	30	30	30	30	30	30
8	36.5	36.5	36.5	36.5	36.5	36.5	36.5
9	40	40	40	40	40	40	40
10	42	42	42	42	42	42	42
11	45	45	45	45	45	45	45
12	45.3	45.3	45.3	45.3	45.3	45.3	45.3
13	49.5	49.5	49.5	49.5	49.5	49.5	49.5
14	69.8	52.5	52.3	42.1	67	27.15	48.3
15	56.2	56.2	56.2	56.2	56.2	56.2	56.2

اذ ان العمود a يمثل قيم المتغير Y بعد تقدير القيمة الشاذة بالطريقة المقترحة الاولى وهكذا بالنسبة الى كل من b ، c ، ....، g على اساس الطرق المقترحة السبعة.

لتحليل بيانات جدول رقم (3) باستخدام انموذج الانحدار التقريبي تم تحديد التقديرات الاولية للمعلمات بنفس الاسلوب السابق والجدول رقم (4) يوضح تلك المقدرات

جدول رقم (4)

يوضح المقدرات الاولية للمعلمات بعد تقدير القيمة الشاذة بالطرق المقترحة

S.M.	Initial Values of Parameters		
	α_0	β_0	γ_0
a	69.9	149.85	0.9560
b	56.3	151.22	0.9435
c	56.3	150.41	0.9437
d	56.3	131.72	0.9488

e	67.2	136.25	0.9574
f	56.3	122.16	0.9517
g	56.3	139.88	0.9465

وبعد تنفيذ برنـامج الانحدار اللاخطـي فـي نظـام SPSS تم التوصل الـى مقدرات المربعات الصغرى للمعلمات عند اختبار فرضية حد الخطأ لحالتي الجمع والضرب وكما موضـح في الجدول رقم (5).

جدول رقم (5)

يوضح مقدرات المربعات الصغرى للمعلمات عند اختبار فرضية حد الخطأ لحالتي الجمع والضرب حسب أنموذج (23) بوجود القيمة الشاذة ومع تقدير القيمة الشاذة بالطرق المقترحة

	Add.					*Mult.*				
Models	α	β	γ	σ^2	R^2	α	β	γ	σ^2	R^2
With Outlier	9650.10	9657.588	0.999	133.03	0.82	65.27	166.128	0.954	0.03697	0.88
a	225.80	232.666	0.995	22.06	0.96	60.58	103.853	0.965	0.01135	0.97
b	104.56	113.032	0.989	6.36	0.98	56.17	101.601	0.961	0.00457	0.99
c	103.96	112.453	0.989	6.39	0.98	56.13	101.580	0.961	0.00455	0.99
d	81.22	90.724	0.984	14.84	0.96	53.40	100.469	0.959	0.00679	0.98
e	188.30	195.420	0.994	17.01	0.96	60.60	103.757	0.965	0.00946	0.98
f	63.01	74.025	0.978	50.93	0.86	48.68	99.271	0.954	0.03096	0.92
g	93.43	102.319	0.987	8.16	0.98	55.11	101.132	0.960	0.00462	0.99

5. الاستنتاجات والتوصيات:

1. تبين من تحليل البيانات أن أنموذج الانحدار التقريبي يعتبر موفق للبيانـات ومـن نتائج جدول رقم (5) نلاحظ انخفاض قيمـة MSE عند فرضية حد الخطأ لحالـة الضرب عن ما هو عليـه فـي حالة الجمع ..

2. بعد توليد قيمة شاذة في بيانات جدول رقم (1) وثم اعادة تحليل البيانـات بوجود القيمـة الشاذة مرة وبعد تقدير القيمة الشاذة بالطرق المقترحة مرة اخرى. نلاحظ انخفاض قيمة MSE وزيادة فـي قيمـة R^2 لحـالتي الجمـع والـضرب ، بعد اجـراء تعديل للقيمـة الـشاذة مـن خـلال تقـديرها بـالطرق المقترحة كانـت نتـائج الطريقتين الثانيـة والثالثـة (b, c) متقاربـة والتـي تـساوي 52.3، 52.5 وجـاءت

مقارنة للقيمة الحقيقية التي تم حذفها وهي 52.9 ، ومن جدول رقم (5) نلاحظ ان الطريقتين الثانية والثالثة اثبتت كفاءتها في تخفيض قيمة MSE وزيادة قيمة R^2 في حالتي الجمع والضرب .

3. من كل ما جاء في اعلاه يمكننا ان نستنتج ان الطرق المقترحة في تقدير القيمة الشاذة قد جاءت بنتائج ايجابية ولقد تم تطبيقها على اكثر من نوع واحد للبيانات ومع استخدام انواع من النماذج اللاخطية مثل (Gompertz, Logistic، Weibull، Richards) في تحليل البيانات. وتبين ان الطريقتين الثانية والثالثة (b،c) هي الافضل في استكشاف وتقدير القيم الشاذة، وان الانموذج Richards ملائم في تحليل البيانات.

4. ضرورة الاهتمام بعملية تقدير القيمة الشاذة في الدراسات والبحوث المستقبلية من دون محاولة بتر تلك القيم وذلك لاهميتها وتأثيرها على دقة النتائج.

5. ان ما اقترح من طرق جديدة لتقدير القيم الشاذة في حالة المتغير الواحد بحاجة الى تطبيق عملي في مجالات اخرى وتطوير ذلك من خلال استخدام نماذج انحدار اللاخطي بمتغيرين.

6. استخدام طرائق اخرى لاكتشاف القيم الشاذة مع تطبيق عملي على بيانات في المجال الطبي او الهندسي لتعين مدى كفاءة وفعالية تلك الطرق.

- المصادر العربية:

1. الجبوري ، شلال حبيب (1990) " اهمية طريقة اكتشاف وتقدير القيمة (القيم) الشاذة في حالة الانحدار الخطي البسيط " مجلة كلية الادارة والاقتصاد العدد الثاني ص 316-335.

2. ناسي ، نبيل (2001) " تقييم كفاءة طرق تقدير القيم الشاذة لنماذج الانحدار " اطروحة دكتوراه- كلية الادارة والاقتصاد- جامعة بغداد.

- المصادر الاجنبية:

3. **Gallant,A.R.(1987), "Nonlinear Statistical Models"; Wiley,New York.**
4. **Jennrich,R.I. (1995), "An Introduction to Computational Statistics-Regression Analysis", Prentice-Hall International ,INC.**
5. **Ratkowsky, D.A. (1983),"Nonlinear Regression Modeling", a Unified Practical Approach, Marcel Dekker, New York.**
6. **Tukey, J.W. (1977), "Exploratory Data Analysis", Addison-Wesley, Reading, MA.**

....

....

....